

Examination of Convection-Allowing Configurations of the WRF Model For the Prediction of Severe Convective Weather: The SPC/NSSL Spring Program 2004

John S. Kain¹, Steven J. Weiss², Jason J. Levit², Michael E. Baldwin¹, and David R. Bright²

¹Cooperative Institute for Mesoscale Meteorological Studies, The University of Oklahoma, and NOAA/OAR/
National Severe Storms Laboratory

²NOAA/NWS/Storm Prediction Center

Accepted for publication in Weather and Forecasting

August 2005

Abstract

Convection-allowing configurations of the WRF model were evaluated during the 2004 SPC/NSSL Spring Program in a simulated severe weather forecasting environment. The utility of the WRF forecasts was assessed in two different ways. First, WRF output was used in the preparation of daily experimental human forecasts for severe weather. These forecasts were compared to corresponding predictions made without access to WRF data to provide a measure of the impact of the experimental data on the human decision-making process. Second, WRF output was compared directly to output from current operational forecast models. Results indicate that human forecasts showed a small, but measurable improvement when forecasters had access to the high-resolution WRF output and, in the mean, the WRF output received higher ratings than the operational Eta model on subjective performance measures related to convective initiation, evolution, and mode. The results suggest that convection-allowing models have the potential to provide a value-added benefit to the traditional guidance package used by severe weather forecasters.

1. Introduction

Co-location of the Storm Prediction Center (SPC) with the National Severe Storms Laboratory (NSSL), the Cooperative Institute for Mesoscale Meteorological Studies (CIMMS), and other agencies in the Norman, OK Weather Center has facilitated considerable interaction and collaboration on a variety of operationally relevant research problems (e.g., Baldwin et al. 2002; Craven et al. 2002; Stensrud and Weiss 2002; Elmore et al. 2003; Kain et al. 2003a; Thompson et al. 2003; Metz et al. 2004; Banacos and Schultz 2005). Over the past five years, the most visible component of this collaboration has been an annual event held during the peak severe convective weather season. This event, known as the Spring Program (Kain et al. 2003b), has attracted a wide cross section of local and visiting forecasters and researchers. The specific emphasis of this program varies each year, but the underlying structure allows forecasters to evaluate new tools or concepts that emanate

from the research community, while immersing research scientists in the challenges, needs, and constraints of the operational forecasting environment. This approach promotes forecast improvements by accelerating the transfer of science and technology into forecast operations at the SPC and by providing researchers with the knowledge to formulate research strategies that will directly benefit operational forecasting.

The SPC is responsible for the prediction of severe convective weather over the contiguous United States on time scales ranging from several hours to three days. To meet these responsibilities, the SPC issues Convective Outlooks for the Day 1, Day 2 and Day 3 periods to highlight regions with enhanced potential for severe thunderstorms (defined as thunderstorms producing hail ≥ 0.75 inch in diameter, wind gusts ≥ 50 knots, thunderstorm induced wind damage, or tornadoes). These outlooks are issued in both categorical (slight, moderate, or high risk) and probabilistic formats, and at set times each day, regardless of the meteorological situation. When the threat level becomes elevated over a mesoscale area in time and space, the SPC may issue a Mesoscale Discussion (MD) product to notify local National Weather Service (NWS) offices that SPC fore-

Corresponding author: Jack Kain, NSSL, 1313 Halley Circle,
Norman, OK 73069

e-mail: jack.kain@noaa.gov phone: 405-579-0776

casters are closely monitoring the area and to articulate the meteorological reasons for their concern. MDs are often issued as a precursor to convective Watches, the cornerstone of SPC forecast products. Watches are issued when conditions become favorable for the development of severe thunderstorms or tornadoes in a specific area over the next 4-7 hours. They are designed to alert a wide variety of users, including NWS and private meteorologists, the public, emergency managers, broadcast media, and aviation interests, of the threatening environmental conditions.

Historically at the SPC, the issuance of both MDs and convective Watches has been driven primarily by observational data (such as 12-hourly upper-air data, hourly surface observations, and higher frequency remotely sensed data from radar and satellite), which are monitored diligently by forecasters. The transition from MD to Watch often is triggered by a specific observation, such as the first sign of deep convective clouds in satellite or radar data. By waiting for such an observation, SPC forecasters undoubtedly increase the accuracy in their placement of convective Watches. The drawback to this approach is that it limits the lead time that forecasters can provide between watch issuance and the development of severe thunderstorms. Over the last several years, the SPC has been making a concerted effort to find ways to increase Watch lead time without sacrificing placement accuracy or increasing Watch size.

Another pressing challenge for the SPC is to improve predictions of convective mode, or morphology. In recent years it has become evident that the type of severe weather that occurs (tornadoes, hail, or damaging winds) is often closely related to the convective mode that storms exhibit (e.g., Snook and Gallus 2004). While many convective storms are best described as mixed-mode or multi-mode systems, SPC forecasters are particularly wary of the emergence of dynamically unique classes of thunderstorms such as supercells and bow echoes because these phenomena are believed to be associated with a disproportionate share of severe weather reports.

Currently, the prediction of the dominant convective mode is based on assessments of key physical properties (e.g., estimates of instability, shear, convective inhibition, and magnitudes of vertical motion) that are difficult to gauge accurately, as well as concepts derived from cloud scale model results and observational studies (e.g., orientation of surface boundaries relative to mean wind and vertical shear vectors - see Weisman et al. 1988, Bluestein and Weisman 2000, and Dial and Racy 2004). If forecasters could find ways to anticipate convective mode more accurately, the specificity and value of all three convective guidance products - Watches, MDs, and Outlooks - would likely increase as well.

One way to address these forecasting challenges is to explore new methods in Numerical Weather Prediction (NWP). At the SPC, both deterministic and ensemble configurations of numerical models are used routinely, but their primary function is to define the larger-scale environment for convective activity, rather than provide specific details about convective storms in time and/or space. In current operational models these details are absent because convection is parameterized as a subgrid-scale process, but in recent years it has become feasible to produce large-domain numerical forecasts without parameterizing deep convection. Specifically, it has become possible to produce forecasts over CONUS-sized domains and grid spacing of about 4 km in a semi-operational setting. The equivalent resolution of these forecasts is best described as a near-convection-resolving compromise - it appears to be fine enough to obviate parameterization of deep convection even though it exaggerates the scale of individual convective clouds.

In spite of this limitation in scale selection, 4 km grid spacing appears to be sufficient to resolve the dominant circulations in organized convective systems (e.g., Weisman et al. 1997). Furthermore, with appropriate initial and lateral-boundary conditions, models with this resolution can have skill in predicting the initiation and organizational mode of convective systems as much as 36-48 h in advance (Fowle and Roebber 2003; Done et al. 2004). Since this may very well be the class of model used in the next generation of NWP, it behooves us to investigate the value of this type of modeling system for severe convective weather forecasting at the SPC.

Such an investigation was the theme of the 2004 SPC/NSSL Spring Program, where three different versions of the Weather Research and Forecast (WRF) model, each configured with ~ 4 km grid spacing and no convective parameterization, were used to predict convective activity over near-CONUS domains each day. Output from these runs was used to generate a daily experimental forecast of severe weather, but only after a baseline had been established by preparing a control forecast using routine observations and model guidance. Specifically, two probabilistic forecasts of severe weather were prepared, over a regional spatial domain and a Watch-like time frame. The first was a control forecast, designed to emulate current operational practice, with data access restricted to operational data streams. The second was the experimental forecast, prepared with access to high-resolution WRF output, after the first forecast was submitted. Differences between these two forecasts were measured to gauge the impact of the high-resolution output. In addition, numerous aspects of the individual high-resolution forecasts were

systematically evaluated and compared to the same characteristics of current operational models.

A key component of the program was the participation of operational SPC forecasters, whose insights and experience impart a real-world severe weather forecasting perspective when assessing the usefulness of high-resolution WRF models. The primary goal of this study is to use data from the Spring Program to assess whether SPC forecasters can make better predictions of severe convective weather when their current data stream of observational and model data is supplemented with output from convection-allowing forecast models. The specific methods used in the 2004 Spring Program are outlined in the next section, followed by a summary of results and a discussion of their implications.

2. Methods used in the 2004 Spring Program

As has been the case in several previous installations of the Spring Program, the 2004 effort had two primary components: 1) experimental human forecasts for severe convective weather and 2) an evaluation of experimental numerical forecast models. Each of these is described below, following a description of the evaluation methods.

a. Subjective Evaluation

A compelling objective of the Spring Program is to facilitate engaging discussion and lively interaction between forecasters and researchers. One of the ways that we promote this activity is through a subjective evaluation process in which all participants become members of a panel of experts. A new panel is constructed each week in the form of a forecast team, consisting of a minimum of one SPC forecaster, one NSSL or CIMMS modeling expert, and one other forecaster or research scientist. On most days in the 2004 Spring Program, there were five or six panel members with a wide variety of backgrounds. Subjective ratings of both human forecasts and model predictions were obtained by means of consensus among all panel members. Achieving consensus was not always easy, but the deliberation process was very effective in soliciting input from all team members. Consensus ratings were assigned on a scale from 0 to 10, with 10 being a superior rating and 0 corresponding to the lowest possible assessment. All ratings were entered on a web-based form, similar to that described in Kain et al. (2003a).

Diversity of viewpoint is essential for a credible subjective evaluation process. For 2004 such diversity was characteristic of the Spring Program. In all, there were about 50 participants over a seven-week period (April 19 - June 4), including contributors from numer-

ous National Oceanographic and Atmospheric Administration (NOAA) research and forecasting organizations, ten major universities, the Air Force Weather Agency, the National Center for Atmospheric Research (NCAR), and international visitors from Canada and Finland (see Appendix). The variety of backgrounds and perspectives in this group was viewed as a key to minimizing the impact of any personal predispositions that could bias subjective assessments.

b. Human forecasts

Morning activities during the 2004 program revolved around preparation of the control and experimental forecasts. The specific forecast product was designed to be a hybrid between the current operational SPC Watch and Outlook products, in that it was issued at scheduled times for fixed time periods (like an Outlook), but was valid for shorter time frames and smaller areas (like a Watch). It consisted of a probabilistic forecast of severe convective weather, including graphic and text components. The graphic portion depicted probability contours for the occurrence of any severe weather. That is, it did not distinguish between different severe weather types such as large hail, damaging winds, and tornadoes, although the attendant discussion alluded to various distinctions, including a prediction of the likelihood of three possible convective modes: discrete cells, quasi-linear systems, and multi-cellular clusters (see the Spring Program Operations Plan at http://www.spc.noaa.gov/exper/Spring_2004/sp04opsplan.pdf for additional details). In addition, the graphic delineated areas where a 10% or greater probability existed for significant severe events (F2 or greater tornado, hail diameter of 2 inches or larger, or wind gusts of 65 knots or greater). The forecast covered a floating regional domain (approximately 14° longitude by 8° latitude, with precise size determined by pre-configured workstation hardware and software display settings) and a six-hour time frame, centered in both time and space on the greatest threat for severe weather.

Threat severity was determined by examining the operational 1300 UTC SPC Day 1 Outlook, observational data, deterministic model forecasts from the Eta (Black 1994) and RUC (Benjamin et al. 2004) models, short-range ensemble forecasts from NCEP (National Centers for Environmental Prediction) (Du et al. 2004), and by consultation with operational SPC forecasters. On most days the forecasts were valid for the 1800-0000 UTC period, in order to capture the time of anticipated afternoon convective initiation. However, on some days when convective development was not expected to occur before late afternoon or evening, the six-hour forecast time period was shifted to 2100-0300 or 0000-

0600 UTC. As part of the forecast process, teams were asked to predict the time of the first severe weather report by selecting a two-hour window within the forecast period. As discussed in the Introduction, the timing of severe convection is a critically important component of SPC severe weather products, especially severe thunderstorm and tornado Watches.

As part of the forecast experimental design, it is important to emphasize that high-resolution WRF-model output was deliberately and uncompromisingly excluded from the control forecast, in order to more directly determine the impact of the high-resolution models on the severe weather forecasting process. Once this first forecast product was completed (two hours was allowed for preparation), high-resolution model output was then introduced. At the same time all updates of real-time observational data were disabled so that the only new information came from the WRF output. The experimental forecast was prepared over the next hour using the high-resolution model data, coupled with the same model and observational data that was used in preparation of the control forecast. These procedures were used to isolate the impact of the high-resolution output in the forecast-preparation process. The control and experimental forecasts were compared to verifying data and to each other in a subjective assessment that took place the next day, then objective verification methods were applied to the entire dataset after the end of the program.

c. Experimental forecast models

The SPC and NSSL formed partnerships with three major modeling centers to ensure the generation of daily high-resolution forecasts for the 2004 program. Specifically, numerical forecasts were produced by NCEP's Environment Modeling Center (EMC; forecast labeled WRF-EMC), NCAR (forecast labeled WRF-NCAR), and the University of Oklahoma's Center for Analysis and Prediction of Storms (CAPS; forecast labeled WRF-CAPS). The WRF-CAPS output was generated at the Pittsburgh Supercomputing Center, while the WRF-EMC and WRF-NCAR forecasts were generated "in-house" at the corresponding modeling centers.

Table 1 summarizes the different model configurations used for these forecasts. The common element in each of these configurations is the horizontal resolution and domain size; each used approximately 4 km grid spacing and a domain covering $\sim 2/3$ or more of the CONUS (Fig. 1). They differed in terms of their numerical dynamic cores, physical parameterizations, and initial conditions. The WRF-EMC forecasts used the Nonhydrostatic Mesoscale Model (NMM) dynamic core (Janjic 2003; Janjic et al. 2004) and a collection of physics routines derived from parameterizations that have been in use for many years in NCEP's operational modeling systems; both the WRF-NCAR and WRF-CAPS configurations used the Advanced Research WRF (ARW) dynamic core (Skamarock et al. 2005) and a set of physical parameterizations derived primarily from ongoing work at NCAR. The WRF-EMC and WRF-

	WRF-EMC	WRF-NCAR	WRF-CAPS
Dynamic Core	NMM	ARW	ARW
Horiz. Grid Spacing (km)	4.5	4.0	4.0
Vertical Levels	35	35	51
PBL/Turb. Param.	MYJ	YSU	YSU
Microphysical Param.	Ferrier	Lin et al.	Lin et al.
Radiation Param. (SW/LW)	GFDL/GFDL	Dudhia/RRTM	Dudhia/RRTM
Initial Conditions	40 km Eta	40 km Eta	40 km Eta + ADAS +Level II Radar

Table 1. Model configurations used for the high resolution forecasts. MYJ: Mellor-Yamada-Janjic (Janjic 2001); YSU: Yonsei University (Noh et al. 2003); Ferrier: Ferrier et al. (2002); Lin et al. (1983); GFDL: Geophysical Fluid Dynamics Laboratory (Tuleya 1994); Dudhia: Dudhia (1989); RRTM: Rapid Radiative Transfer Model (Mlawer et al. 1997; Iacono et al. 2000); ADAS: ARPS (Advanced Regional Prediction System) Data Assimilation System (Brewster 1996).

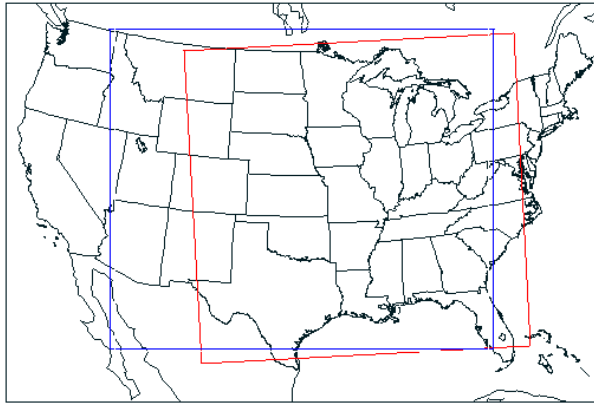


Fig. 1. Model domains for the WRF-EMC (right inset) and the WRF-NCAR and WRF-CAPS configurations (left inset, identical domains)

NCAR forecasts used a "cold-start" initialization, simply interpolating data from the Eta model's 0000 UTC analyses (standard Eta output with 40 km grid spacing) to the high-resolution grids. The WRF-CAPS used the Eta analysis as a first guess but modified the Eta fields using an experimental data assimilation procedure that incorporated hydrometeor fields derived from Level II radar data (Brewster 1996; see Crum et al. 1993 for description of Level II data). All forecasts used the 0000 UTC Eta forecast for lateral boundary conditions.

The radar assimilation used in the WRF-CAPS runs led to realistic rainfall patterns during the first 1–3 h of daily forecasts, but these forecasts were consistently and substantially inferior during the next-afternoon/evening evaluation period. After the end of the program, this poor performance was linked to an error in the specification of lateral boundary conditions (D. Weber, CAPS, Norman, OK, 2004, personal communication). Since the impact of this error was so detrimental as to preclude meaningful comparison with the other model forecasts, evaluation of the WRF-CAPS forecasts is not discussed further in this paper.

This configuration was the only one to exhibit obvious major deficiencies to Spring Program participants, but it is important to emphasize that each of the high-resolution WRF configurations was experimental and had not been subjected to extensive testing prior to the start of the program. Over the course of the program scientists at NCAR became concerned that the WRF-NCAR forecasts "seemed to be overly aggressive in producing boundary layer convection and small-scale convective cells" (M. Weisman, NCAR, Boulder, CO, 2004, personal communication). After the program was com-

pleted they attributed this behavior to a lack of horizontal numerical diffusion, which had been disabled due to an erroneous parameter setting. They later restored the numerical diffusion parameterization and re-ran many of the Spring Program simulations. The new simulations were judged to be different enough to conclude that the wayward parameter setting could have impacted subjective verification ratings. Since it was not possible to repeat the subjective verification with the new simulations, the mean subjective ratings earned by the WRF-NCAR forecasts are not shown here. Scientists at EMC also expressed some concern about their configuration at the start of the experiment, particularly a perceived high precipitation bias (M. Pyle, EMC, Camp Springs, MD, 2004, personal communication). Nonetheless, they agreed to maintain a stable configuration through the end of the program and to allow their subjective verification results to be published. All configurations of the WRF model are still under development.

Because the full high-resolution output datasets from these configurations were very large (~2 Gb per output time), only a subset of the full output was transferred to the SPC and ingested into the data stream for the program, focusing on selected output fields that are commonly examined by severe weather forecasters. Specific output fields included instantaneous precipitation rate, one-hourly and three-hourly accumulated rainfall, low-level wind and moisture fields (including their derivatives such as mass and moisture convergence), CAPE (Convective Available Potential Energy), CIN (Convective INhibition), and several vertical shear and storm-relative helicity parameters. Model verification ratings were based primarily on the hourly accumulated rainfall field and its comparison with an hourly radar mosaic base-reflectivity field (displayed as the maximum reflectivity at each pixel during the previous hour). Ideally, a more direct comparison with radar observations could be made by computing an equivalent reflectivity field from instantaneous model hydrometeor fields, but the necessary model output and post-processing algorithms for equivalent reflectivity were not available for all models at the start of the program. One-hour precipitation fields proved to be quite adequate for identifying the timing and location of convective initiation. Assessing the general characteristics of convective evolution was somewhat more challenging since it required a mental integration over the 6 h forecast period. The third verification category, convective mode, proved to be much more challenging, as it required one to consider details of the variations in mode across the domain and over the 6 h time period. Furthermore, the convective mode assessment was also applied to output from the Eta model, which typically provides little, if any, information about the internal structure of convective sys-

tems (Baldwin and Wandishin 2002). Concerns about the mode evaluation are discussed in more detail in the next section. All model output was displayed and verified on the respective model's native grid.

Because of the substantial time required to integrate the WRF models and generate the high-resolution output, we were forced to initialize the experimental models with 0000 UTC data. However, it should be noted that forecasts from the 1200 UTC initializations of the operational models (Eta and RUC) were used both for forecast guidance and for subjective comparison. The 1200 UTC runs of the Eta and RUC models were used for two reasons: 1) the desire to emulate operational routines in preparing the control forecast - SPC forecasters typically focus on the most recently updated model guidance in preparing forecasts operationally, and 2) operational RUC forecasts are only 12 h in length and guidance was needed for the afternoon to evening time period. Although the latest updates of model forecasts are not always the most skillful (e.g., see Kain et al. 2003a), it is recognized that this approach may handicap

the high-resolution models in a direct comparison with the 1200 UTC RUC and Eta runs. Note that, as with the higher resolution models, output from the Eta and RUC models was displayed on native model grids (spacing of 12 and 20 km, respectively).

3. Results

Results from the 2004 program are first illustrated by using one day's forecast as an example, followed by an overall assessment of human forecasts, and an overview of all model forecasts.

a. Example of a severe weather forecast

High-resolution model guidance had a significant positive impact on human forecasters on 28 May. As forecast teams assessed the meteorological scenario on this day, they noted that an upper level ridge was in place over the central and northern Plains while an embedded short-wave trough was passing over north-

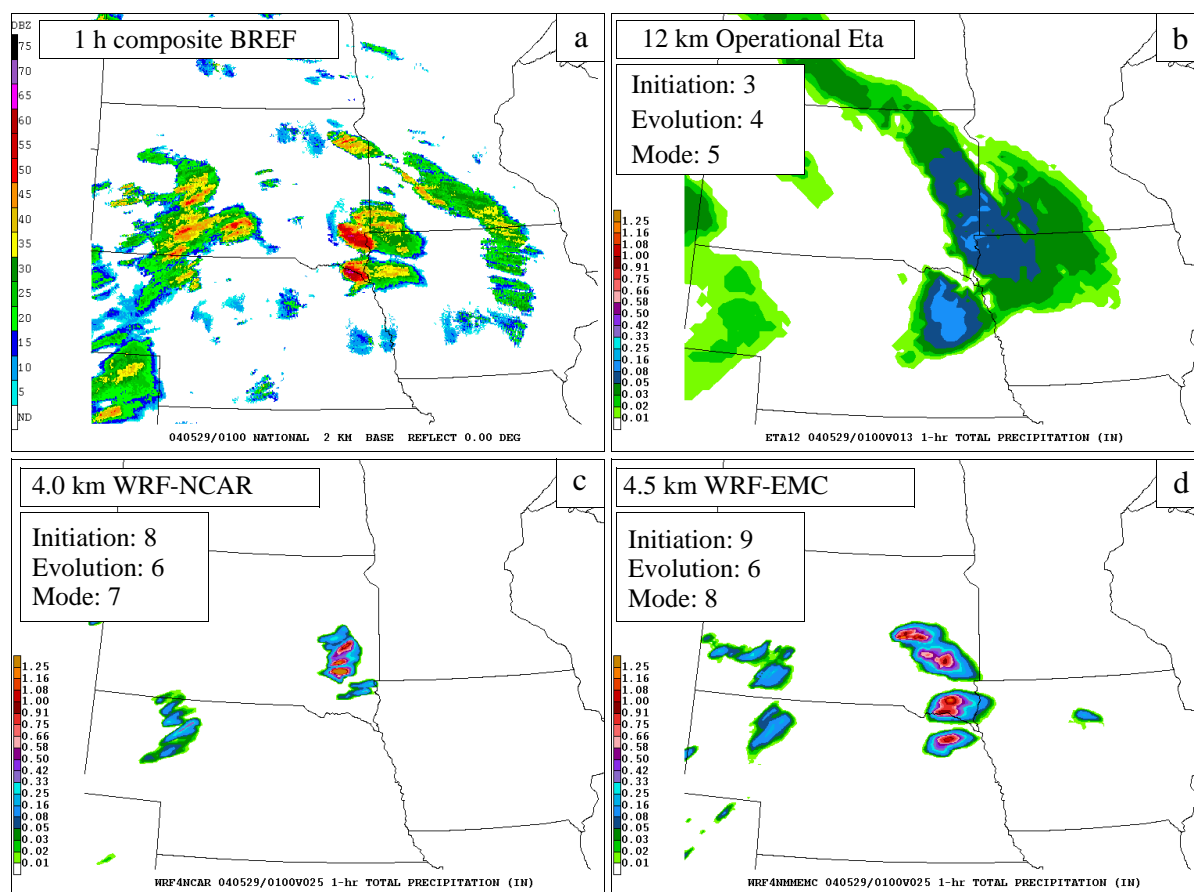


Fig. 2. Verifying radar data and model forecasts valid 0100 UTC 29 May 2004. Time composite of base reflectivity (a) is derived as maximum reflectivity at each pixel over the previous hour. Model forecasts show accumulated precipitation over the previous hour. Subjective verification ratings are indicated in the upper left of the model-forecast frames.

eastern Wyoming (not shown). This trough was expected to move over the Dakotas by late in the day and, in conjunction with a lee trough and associated warm front at the surface, to trigger convection in this region. Wind fields and surface-based instability were judged sufficient to support severe thunderstorms, including isolated supercells.

Observational data and operational model guidance suggested that precipitation would develop over this region between 2100 and 0000 UTC and move eastward with the prevailing flow, thus the forecast team focused on the 2100 - 0300 UTC time frame and a regional domain centered on Sioux Falls, South Dakota for control and experimental forecasts. Since the 1200 UTC RUC guidance was available only through 0000 UTC, forecast teams relied most heavily on the Eta model for deterministic guidance in preparation of the control forecast. Between 0000 and 0100 UTC the Eta model predicted a broad swath of precipitation along the central Iowa-Minnesota border, with a lobe extending southwestward into northeastern Nebraska and an additional extension along the warm front towards the northwest (Fig. 2b). Based largely on this coverage pattern, the forecast team outlined a large area with 15% probability of severe weather in the control forecast, extending from west-central North Dakota southeastward into South Dakota and encompassing parts of Minnesota, Iowa, and Nebraska (Fig. 3a).

When the high-resolution model output was made available, forecast teams were presented with a different scenario. In particular, the WRF-EMC and the WRF-NCAR runs developed intense convection over a much smaller area, concentrated in southeastern South Dakota (Figs. 2c and d), with little precipitation elsewhere. Forecasters developed confidence in this signal for a relatively isolated region of intense convection, inspired by the consistency between these two model forecasts, the reasonable behavior of these models during the first ~ 15h of integration, and the consistent evolution of other fields in the model guidance. Their response was to reduce the areal coverage of the 15% probability contour substantially and add an area with 25% probability over southeastern South Dakota (Fig. 3b).

When severe weather reports were examined the next working day, it was quite obvious that the high-resolution models had a favorable impact (note the location of concentrated severe weather reports in Figs. 3a, b). Since these forecasts were made on a Friday, they were evaluated by a new forecast team on Monday morning. The control forecast received 5 points out of 10, earning credit for encompassing all reports within the 15% contour, but penalty points for the large area farther to the northwest where no reports were received. By comparison, the experimental forecast that benefited from the

high-resolution numerical guidance was given 8 points out of 10, as high as any forecast during the entire program. The improvement on this day (3 points) was higher than on any other day on which both high-resolution models were available.

All NWP forecasts from this Friday were also evaluated by the new forecast team (note ratings in the upper left hand corners of Fig. 2b, c, d). Although Fig. 2 shows only one of the six hourly frames that were used to evaluate the models, it provides a good sense of the differences between the Eta, WRF-EMC, and WRF-NCAR forecasts and their relative correspondence with radar data. Other model forecasts were also evaluated on this day, but for the sake of brevity, herein we limit

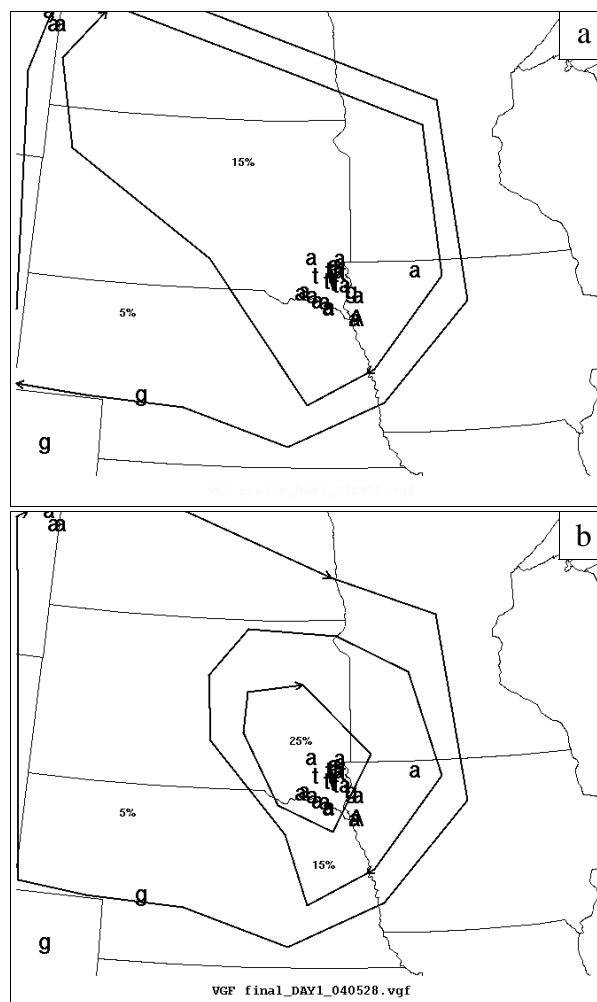


Fig. 3. a) control and b) experimental severe weather forecasts valid 2100 UTC 28 May - 0300 UTC 29 May 2004. The outermost contour encloses the =5% probability area, while additional contours indicate =15% and =25% probability. Severe weather reports from this period are denoted by the letters t (tornado), a (hail = 0.75 in.), and g (wind gusts = 50 kts).

discussion to the highest rated WRF configurations and the "benchmark" Eta model.

In terms of convective initiation, the WRF-EMC forecast received an exceptionally high score of 9 out of 10, showing excellent correspondence with observations in both timing and location of severe storm development in southeastern South Dakota. The WRF-NCAR was only one point lower, as it was 1-2 h late in activation, but excellent in placement of intense convection. The Eta model was penalized quite heavily by the forecast team (3 out of 10) because it activated parameterized convection too early and over much too broad of an area, with no focal point for more intense activity.

In evaluating convective evolution, forecast teams were instructed to focus on direction and speed of system movement, areal coverage, and configuration and orientation of mesoscale features. The nondescript structure of the Eta precipitation field left a negative impression with forecast teams in this category as well. It received a rating of 4 out of 10, with archived comments indicating credit for predicting the direction and speed of movement quite well, but penalty points for too much coverage and "obscured configuration of mesoscale structures." The two WRF configurations fared somewhat better, both receiving a 6 in the evolution category. Both of these forecasts were credited with forecasting the location, movement, configuration and orientation well, but they were criticized for underpredicting the areal coverage of weaker convection.

The two WRF forecasts received high ratings for convective mode (8 for the WRF-EMC and 7 for the WRF-NCAR) because they correctly predicted intense isolated convective cells where most of the severe reports occurred, with a slight penalty for missing some non-severe multi-cellular convection that developed elsewhere in the forecast domain. The Eta model earned a 5 for convective mode, producing a "blobbish" precipitation field that was categorized as multi-cellular by the process of elimination, i.e., because neither quasi-linear structures nor isolated cells could be discerned.

b. Human forecasts: Areal coverage of severe convection

On most days, forecast teams made relatively minor updates to the control forecasts when they issued their experimental counterparts (in contrast to the significant adjustment shown in Fig. 3). This is consistent with routine practice at the SPC: operational forecasters tend to make only incremental changes when Outlook updates are issued unless they discover compelling evidence that major modifications are needed. This approach stems partly from the need of Outlook users for a high degree of consistency and continuity from one

forecast to the next, but it is also prudent because every existing forecast contains a certain amount of inertia, having been systematically assembled from a large body of evidence, including observational surface and upper air data, multiple derived convective parameters, satellite and radar imagery, operational mesoscale and SREF modeling systems, and forecaster experience. During the Spring Program, the control forecasts carried the same weight of supporting evidence, seemingly insulating them from major adjustments. Furthermore, significant changes were difficult to justify because performance characteristics and systematic biases of the particular WRF configurations were relatively unknown to the new forecast teams that were assembled each week. Consequently, forecasters tended to proceed with caution in formulating experimental forecasts, taking measured steps in response to the new high-resolution guidance.

Presumably, these measured steps were taken solely in response to one new category of information - the convection-allowing model forecasts. The program was designed to ensure this by withholding other relevant information, such as observational updates. Yet, some factors could not be excluded. For example, it is possible that some forecasters "tweaked" their forecasts simply because they had more time to think about the meteorological scenario. Such factors cannot be ruled out, but it seems safe to assume that the three different scenarios presented by the high-resolution models dominated the thought processes of the forecast teams and exerted the strongest influence on any decisions to modify the control forecasts.

Control and experimental forecasts were verified with one subjective and two objective measures, using severe weather reports as ground truth. The subjective verification was based on next-day panel evaluations of the accuracy and usefulness of each forecast to potential customers, such as NWS forecasters and emergency managers, focusing on areas with greater observed coverage of severe reports and/or higher forecast probabilities within the regional domain. The evaluation teams also had access to radar signature and severe weather warning information, which was used to supplement the severe reports in regions where low population might affect the number of ground-truth reports of severe weather. Particular attention was given to the skill of the experimental forecast relative to the control (e.g., was it better, worse, or similar in accuracy and usefulness?). In this way, although the panel members varied from week to week and the raw rating numbers were not uniformly calibrated, the difference between the forecast ratings could be used to assess relative skill.

Objective measures of forecast skill, verified exclusively against local severe weather reports, were com-

puted using the Brier Score (Brier 1950) and the area under the Relative Operating Characteristic (ROC) curve (Mason 1982). The Brier Score is commonly used to verify probabilistic forecasts and ranges from 0 to 1, with 0 being perfect (lower scores indicate better forecasts). The ROC is also useful for verifying probabilistic forecasts and their ability to discriminate occurrences from non-occurrences. If the area under the ROC-curve is integrated, values range from a perfect score of 1 to a useless value of <0.5 , with an area of >0.7 considered to represent reasonable discriminating capability. Severe weather reports and forecast probabilities were both mapped to an 80 km grid for objective verification, roughly consistent with the concept of detecting severe weather within ~ 25 miles of a point (see Brooks et al. 2003 for details on the mapping strategy).

Results from the various verification measures are summarized in Table 2, based on the 29 days on which both control and experimental forecasts were issued. Each of the three approaches provides a more favorable overall result for the experimental forecast. The difference in mean subjective ratings is statistically significant, but the differences are very small for the objective measures and not likely to be statistically significant. All three measures indicate that the experimental forecast was more skillful on about half of the days. As for those days when the experimental forecast was not categorized as more skillful, the subjective approach yields a relatively large number of tie scores, while the objective metrics generate comparatively few ties and suggest that the experimental forecast was less skillful than the control on almost half of the days. The nature of this discrepancy is not entirely clear, but it should be noted that many of the negative changes in the objective scores are quite small, perhaps proportionately less than the smallest increment available in subjective ratings (1 point).

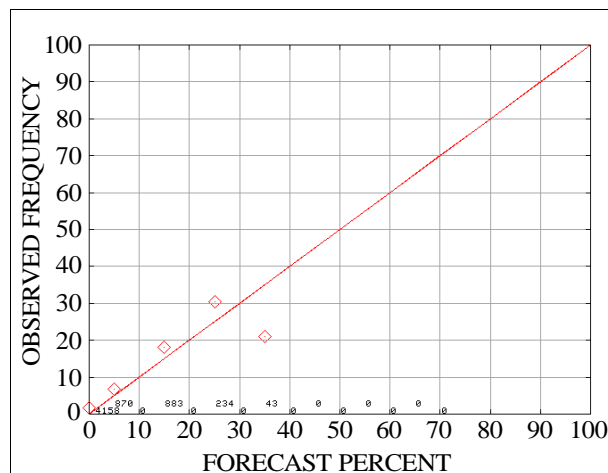


Fig. 4. Reliability diagram for the experimental human forecasts.

Reliability diagrams (Wilks 1995) for the control and experimental forecasts were nearly identical, and only the experimental forecast diagram is shown (Fig. 4). This diagram indicates excellent reliability for all probability values below 35%. For these values the appearance of slight under-forecasting is largely an artifact of the allowable forecast probability intervals (5%, 15%, 25%, etc.), such that a 25% value actually represents all probabilities in the 25-34% range. The slight over-forecast bias evident in 35% probability forecasts may reflect a very small sample size, as these forecasts included only 43 total grid blocks.

In general, these subjective and objective verification measures are consistent in indicating that the high-resolution model data had a small positive impact on the experimental severe weather forecasts. The subjective scores seem to provide a more favorable relative assessment, perhaps reflecting the consideration of additional data sources (radar signatures and warning information) in the subjective evaluation process.

	Subjective Ratings	Brier Scores	ROC Curve Area
Control Forecast	5.3 (mean)	0.0514	0.778
Experimental Forecast	5.9 (mean)	0.0511	0.794
Days with Forecast Improvement	15	14	14
Days with Forecast Degradation	4	12	11
Days with Unchanged Ratings/Scores	10	3	4

Table 2. Summary of subjective ratings and objective verification scores for the control and experimental human forecasts.

c. Human forecasts: Timing of first severe report

The ability of the forecast teams to predict the timing of the first severe report, within a two-hour time window, was also explored. The correct time window was selected in the control forecast on 11 days. The forecast team chose to make an adjustment to the control-forecast window in the second (experimental) forecast only 3 times (all 3 were on days when the initial window turned out to be incorrect), but all of these adjustments were in the right direction and two of them captured the first severe report. These results are consistent with the small improvements noted above. At the same time, when compared to the total number of forecasts, they give some indication of just how difficult it can be to predict convective initiation accurately. Neglecting 4 days on which there were either no severe reports within the forecast domain or the first reported event occurred before the experimental forecast was issued, these numbers reveal that the 2-hour time window was correctly placed just over 50% of the time in the experimental forecasts (13 out of 25 forecasts). These findings suggest that specific prediction of the onset of severe weather remains very challenging, as

details of convective initiation and subsequent intensification are dependent on mesoscale and storm-scale processes that are poorly resolved by observational data and not well understood by the meteorological community.

d. Overall Assessment of Model Forecasts

Although up to eight different model forecasts were verified on some days, discussion herein focuses on a comparison of output from the Eta model (as an operational benchmark) and the WRF-EMC (as the most robust high-resolution configuration). The mean ratings over this period favored the WRF-EMC output in every category (Fig. 5). Differences between the WRF-EMC and Eta ratings are statistically significant in the initiation and mode categories and nearly so in evolution.

The relatively high initiation ratings earned by the convection-allowing configuration are somewhat surprising. Previous studies have suggested that convective initiation should be delayed in model simulations with coarsely, but explicitly resolved convection compared to analogous runs using parameterized convection or higher resolution (e.g., Molinari and Dudek 1986;

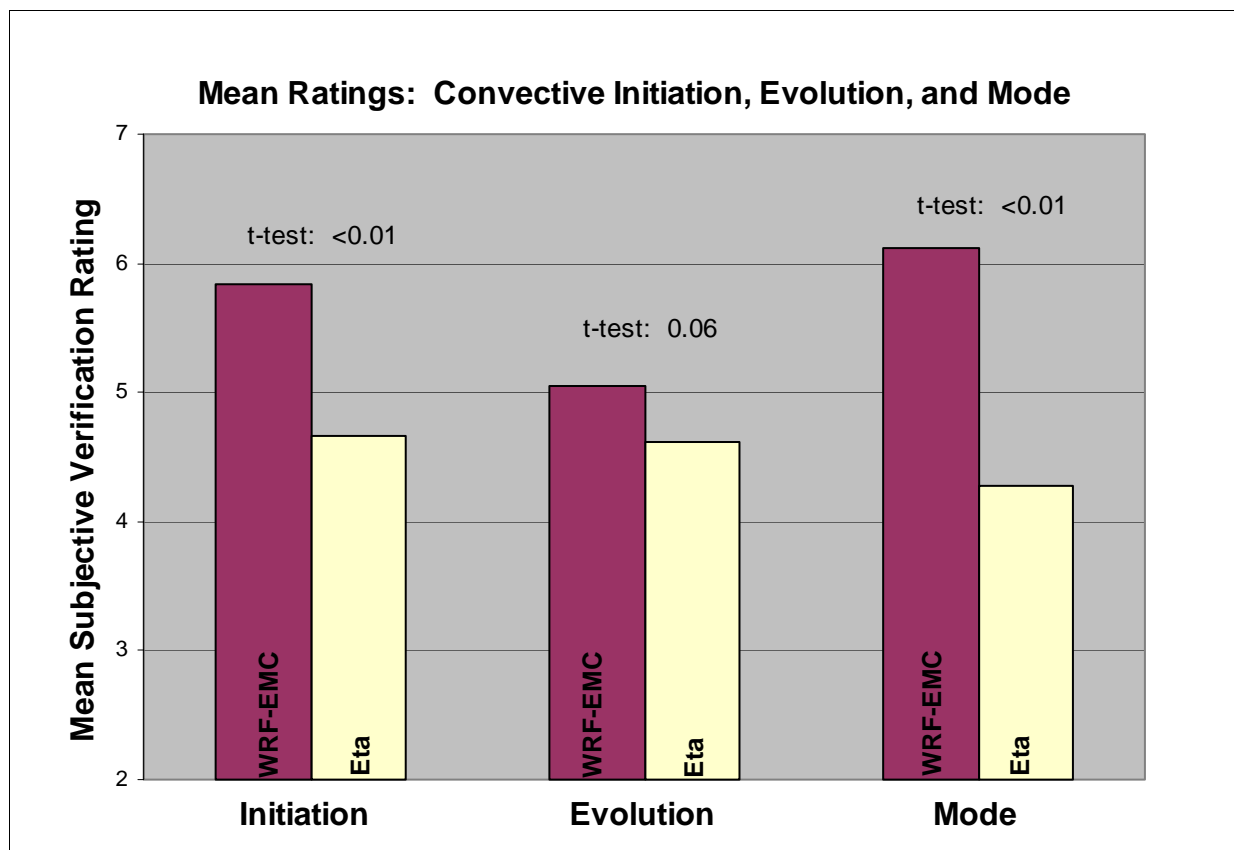


Fig. 5. Mean subjective verification ratings for the operational Eta model and the WRF-EMC configuration, for categories of convective initiation, evolution, and mode. Paired t-test scores are listed above each category to indicate statistical significance.

Weisman et al. 1997). Thus, one might have expected initiation in the WRF-EMC forecasts to occur later and downstream relative to the initial parameterized convection from the Eta model. Such a tendency was not evident during the program, but one should use caution in comparing this result to previous studies. In this case, the high-resolution (WRF-EMC) and low-resolution (Eta) models were configured quite differently and many factors could have influenced the timing and location of convective initiation. Forecast teams often noted that low-level moisture and CAPE values in the WRF-EMC output seemed higher than corresponding fields from the Eta model. If this perceived bias was real, it may have helped the WRF-EMC model initiate convection earlier than it otherwise would have.

Since convective initiation is, in effect, the first stage of evolution, these two categories are inherently related. In terms of the subjective ratings, this association emerges as a modest positive correlation between initiation and evolution for both models (Table 3). Convective mode ratings also show a positive correlation with initiation for the WRF-EMC model, but the correlation between these two parameters is actually negative for the Eta model. Similarly, mode ratings show little or no correlation with evolution ratings for the Eta model, but a weak to moderate positive correlation for the WRF-EMC. Thus, ratings in all three categories are correlated when the high-resolution output is examined, but convective mode assessments seem to be somehow different with the Eta model.

Of course, the character of the Eta model's precipitation fields is distinctly different from that of convection-allowing models. Convective precipitation is parameterized in the Eta model. Precipitation patterns in this model rarely (if ever) contain a level of internal structure that is comparable to corresponding radar images (e.g., Baldwin and Wandishin 2002). Instead, the Eta often produces relatively amorphous, low amplitude "blobs" of convective precipitation. During the program, when these blobs had an aspect ratio of more than 5:1 and a length scale of at least 100 km, they were

labeled linear convective systems; when their dimensions were smaller than 100 km or so, they were labeled isolated cells; otherwise, they were classified as multi-cell clusters. In the absence of internal structure, the convective mode predicted by the Eta model was implied simply by the shape and size of contiguous precipitation areas. In contrast, the 4-km WRF output typically contained many high amplitude, small-scale structures - some isolated and some embedded within larger precipitation features - similar to the patterns associated with real convective systems. Grid spacing of 4 km is not sufficient to resolve convective-scale features well, but it does allow for the development of some detailed internal structures within convective systems. With the 4 km forecasts, mode interpretation often focused as much on the smaller-scale details within contiguous precipitation regions as on the shape of the regions' outline. But even at 4 km, precipitation patterns are a poor discriminator of convective mode. While the difference in convective mode ratings between the Eta and WRF-EMC is statistically significant, interpretation of these ratings is complicated by the inherent dissimilarity in the character of precipitation fields produced by these two models. The initiation and evolution categories are much less ambiguous and should be weighted more heavily in an overall evaluation of model performance.

In spite of the difficulties that inevitably arise, systematic comparisons of forecasts from state-of-the-art convection-allowing and convection-parameterizing models are essential because the meteorological community is trying to decide where to dedicate NWP resources that will be available in coming years. Until suitable objective verification metrics are developed for this specific application, subjective assessments such as this one will provide the best guidance in deciding which path to follow. The subjective ratings shown here, especially those in the initiation and evolution categories, reveal that participants in the 2004 Spring Program perceived a convection-allowing configuration of the WRF model as a better predictor of severe convec-

CORRELATION

	Initiation vs. Evolution	Initiation vs. Mode	Mode vs. Evolution
12 km Eta	0.71	-0.45	-0.17
4.5 km WRF-EMC	0.72	0.55	0.64

Table 3. Correlation between ratings in different categories for the Eta and WRF-EMC models.

tive activity than the operational Eta model, on average. This suggests that similarly configured high-resolution models have the potential to provide unique and valuable information for severe weather forecasters, though it remains to be determined whether the considerably higher cost of the high-resolution forecasts is commensurate with the added value.

4. Discussion

Output from the high-resolution WRF models was similar to operational forecast models (e.g., the Eta and RUC models) in that it showed overall good correspondence with observed convective activity on some days but poor correspondence on other days. Undoubtedly, initial and boundary conditions, which were derived from the 0000 UTC Eta model forecast, played a large role in the overall quality of the forecasts. Yet, on nearly every day the precipitation fields from the WRF forecasts had a much more realistic appearance than the relatively smooth, less structured output that forecasters are accustomed to seeing from operational models. Because they resembled radar reflectivity patterns more closely, the precipitation forecasts looked more believable. Forecasters were naturally intrigued by and attracted to the high-resolution output. With each new group of forecasters, this attraction seemed to inspire an initial false sense of confidence in the forecasts, but each week during the program forecast teams quickly adopted a more objective approach after critically examining WRF forecasts and comparing with observations (and operational model forecasts) while the forecast scenario was still fresh in their mind.

This process was important because the introduction of any new model guidance into the forecast decision-making process requires forecaster determination of the likelihood that the guidance will provide useful information. Typically, the first step towards building confidence in a given model forecast is to determine the extent to which the model is "on the right track." For example, during the program forecast teams routinely examined numerous forecast fields that are relevant to convective forecasting (e.g. CAPE, CIN, surface temperature, dewpoint, and wind fields, vertical shear, etc.) before deciding how much confidence to place in a given model forecast. These fields were not only useful for comparison with observed counterparts to determine if the first 12-15 hours of the model forecast appeared reasonable, they were also instrumental in linking the model precipitation to appropriate physical mechanisms.

Another important factor in weighing the significance of different guidance products involves application of knowledge about model configuration and

physics (e.g., Baldwin et al. 2002, Kain et al 2003c), especially as these characteristics relate to known model performance characteristics. But forecasters had relatively little knowledge about the performance characteristics of the high-resolution models prior to the start of the Spring Program. The combination of thorough next-day examination of model output and direct interactions with modeling experts accelerated the learning process. Our experience during the program suggests that specialized severe weather forecasters working closely with model developers and research scientists can combine their complementary knowledge bases and begin to identify ways of applying experimental high resolution model forecasts to address operational forecasting needs within relatively short periods of time. Thus, continued collaboration and dialogue between the operational forecasting and model development communities is essential in order to expedite the development of operationally relevant NWP systems and subsequent transfer of these systems from research to operations.

5. Summary

The 2004 SPC/NSSL Spring Program focused on testing experimental, convection-allowing configurations of the WRF model to determine if they can provide new and unique information from the perspective of operational severe weather forecasters, and, if so, whether this information can be used to improve severe weather forecasts. The program was conducted over a seven-week period during the peak severe weather season. Each week, a new team was rotated into the program, with all teams consisting of five or six people - a mix of forecasters, model developers, and research scientists. Their daily tasks included preparation of experimental forecasts, verification of previous forecasts, and detailed examination and verification of model output.

Three different configurations of the WRF model were utilized during the program, all of which used a grid spacing of about 4 km, and a domain covering most of the CONUS east of the Rockies (Fig. 1). Two runs used an initial condition derived from simple interpolation of an Eta model analysis, while the third enhanced this first-guess field by assimilating radar data. The model forecasts tested different dynamic cores and physical parameterizations (Table 1). All three configurations should be considered experimental since none had been subjected to a long period of "fine-tuning" to ensure that all of the model dynamics, physics, and parameterization schemes were well calibrated. Forecasts from these high-resolution models were compared to those from the Eta model to provide a relative measure of performance compared to an operational benchmark.

Experimental forecasts and model evaluations were based on limited criteria related to forecasting responsibilities of the SPC. In particular, they focused on specific topics of convective initiation, evolution, and mode that directly impact the issuance of SPC Mesoscale Discussion and Severe Thunderstorm/Tornado Watch products. Thus, experimental forecasts were made over limited regional domains, centered on the area of greatest threat for severe weather, and over 6 h periods, similar to the period of a convective Watch. Model evaluation took place over the same temporal and spatial domain.

In general, post-program analysis of both experimental human forecasts and model guidance suggests that the high-resolution output had a small, but positive impact in the severe weather forecasting environment. Based on mean subjective verification ratings assigned by the weekly forecast teams, the most robust WRF configuration earned higher ratings than the Eta model in all three categories of convective initiation, evolution, and mode. Furthermore, experimental probabilistic forecasts issued by the weekly teams after they had examined high-resolution output were rated as more skillful, on average, than control forecasts that were prepared using only routine operational data for guidance. The small advantage for the experimental forecasts was corroborated by objective verification measures.

Collectively, results from the Spring Program provide compelling evidence that the tested experimental models provided added value for forecasters of severe weather. But these results do not preclude the possibility that other numerical guidance products could offer similar value. The difference in computational cost of running at 4 km vs. the current 12 km grid spacing of the Eta model is at least a factor of 27. It is possible that a cheaper alternative could provide the same benefit. However, increasing resolution to that afforded by 4-5 km grid spacing seems particularly attractive because it obviates the need for parameterization of deep convection, which is viewed by many as the Achilles heel of current operational NWP models.

The 2004 Spring Program provided valuable and unique feedback to model developers as they continue to test the WRF model and optimize its performance. It clearly benefited the operational forecasting community as well, introducing them to state-of-the-art developments from leading research centers. Furthermore, the program had more widespread, if less tangible benefits. On an individual level, many participants reported that the most rewarding aspect of their involvement was the unique and lively interaction between forecasters and researchers. Working relationships were forged and mutual respect was earned, perhaps because of the challenging demands that were placed collectively on all

participants. Forecasters and researchers alike were required to make consensus predictions in the face of uncertainty, to critically examine their decisions as verifying data became available (often a humbling experience), and to arrive at a consensus on numerous other subjective assessments. Through this process, forecasters learned more about the scientific basis for new research developments and researchers developed an appreciation for the uncertainty and operational constraints associated with the daily challenges faced by SPC forecasters. The program "greases the skids" for the transfer of science and technology into forecast operations and it makes participating research scientists better equipped to formulate and conduct operationally relevant research. For those participants who have teaching responsibilities, the program also provides them with the knowledge to bring "real-world" relevance into the classroom. Several external participants in the 2004 Spring Program indicated that their experience in the program would have a direct impact on the content of courses that they teach at major universities. Thus, the influence of the Spring Program is far reaching, touching and uniting the areas of operational forecasting, scientific research, and university instruction.

Acknowledgements

We are very grateful to numerous individuals at CAPS, EMC, and NCAR for their valuable collaboration and extraordinary efforts to generate and disseminate high-resolution model output on a daily basis. In particular, we thank Kevin Thomas, Dan Weber, and Kelvin Droegeemeier from CAPS; Matt Pyle, Zavisa Janjic, and Geoff DiMego from EMC; and Wei Wang, Morris Weisman, and Joe Klemp from NCAR. We thank Adam French, a student from Valparaiso University who visited the SPC under the ORISE program and compiled much of the data discussed herein. In addition, we thank all participants in the 2004 Spring Program (see Appendix). This program would not be possible without the technical support of many people at the NSSL/SPC/CIMMS facility in Norman, especially Phillip Bothwell, Gregg Grosshans, Jay Liang, and Doug Rhue from the SPC; and Brett Morrow and Steve Fletcher from Indus/NSSL. Linda Crank (SAIC/ SPC), Tracy Reinke (CIMMS/NSSL), and Peggy Stogsdill (SPC) are deeply appreciated for making complex travel arrangements as simple as possible. Finally, we would like to thank COMET, an organization that has supported this and previous collaborative activities at the Oklahoma Weather Center. The 2004 Spring Program was partially supported by COMET Cooperative Project S03-44678 and COMET Partner's Project S04-44693. Additional support for authors Kain and Baldwin came from

NOAA-OU Cooperative Agreement NA17RJ1227. NSSL provided partial support for travel expenses. This manuscript benefited from thorough reviews by Phillip Spencer of CIMMS/NSSL and three anonymous reviewers.

REFERENCES

- Baldwin, M. E., J. S. Kain, and M. P. Kay, 2002: Properties of the convection scheme in NCEP's Eta model that affect forecast sounding interpretation. *Wea. Forecasting*, **17**, 1063-1079.
- Baldwin, M. E., and M. S. Wandishin, 2002: Determining the resolved spatial scales of Eta model precipitation forecasts. *Preprints, 15th Conference on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., 85-88.
- Banacos, P. C., and D. M. Schultz, 2005: The use of moisture flux convergence in forecasting convective initiation: Historical and operational perspectives. *Wea. Forecasting*, **20**, 351-366.
- Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Smirnova, and R. Bleck, 2004: Mesoscale weather prediction with the RUC hybrid isentropic-terrain-following coordinate model. *Mon. Wea. Rev.*, **132**, 473-494.
- Black, T. L. 1994: The New NMC Mesoscale Eta Model: Description and Forecast Examples. *Wea. Forecasting*, **9**, 265-284.
- Bluestein, H. B. and M. L. Weisman, 2000: The interaction of numerically simulated supercells initiated along lines. *Mon. Wea. Rev.*, **128**, 3128-3149.
- Brewster, K., 1996: Application of a Bratseth analysis scheme including Doppler radar data. *Preprints, 15th Conf. on Weather Analysis and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., 92-95.
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1-3.
- Brooks H. E., C. A. Doswell III, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626-640.
- Craven, J. P., R. E. Jewell, and H. E. Brooks, 2002: Comparison between observed convective cloud-base heights and lifting condensation level for two different lifted parcels. *Wea. Forecasting*, **17**, 885-890.
- Crum, T. D., R. L. Alberty, and D. W. Burgess, 1993: Recording, archiving, and using WSR-88D data. *Bull. Amer. Meteor. Soc.*, **74**, 645-653.
- Dial, G. L. and J. P. Racy, 2004: Forecasting short term convective mode and evolution for severe storms initiated along synoptic boundaries. *Preprints, 22nd Conference on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., CD-ROM, paper 11A.2
- Done, J., C. A. Davis, and M. L. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) model. *Atmos. Sci. Let.*, published online 8 September 2004, p. 110-117.
- Du, J., J. McQueen, G. J. DiMego, T. L. Black, H. Juang, E. Rogers, B. S. Ferrier, B. Zhou, Z. Toth, and S. Tracton, 2004: The NOAA/NWS/NCEP Short Range Ensemble Forecast (SREF) system: Evaluation of an Initial Condition vs Multiple Model Physics Ensemble Approach. *Preprints, 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., CD-ROM, 21.3
- Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077-3107.
- Elmore, K. L., S. J. Weiss, and P. C. Banacos, 2003: Operational ensemble cloud model forecasts: Some preliminary results. *Wea. Forecasting*, **18**, 953-964.
- Ferrier, B. S., Y. Jin, Y. Lin, T. Black, E. Rogers, and G. DiMego, 2002: Implementation of a new grid-scale cloud and precipitation scheme in the NCEP Eta model. *Preprints, 15th Conf. on Numerical Weather Prediction*, Amer. Meteor. Soc., San Antonio, TX, 280-283.
- Fowle, M. A., and P. J. Roebber, 2003: Short-range (0-48 h) numerical prediction of convective occurrence, mode and location. *Wea. Forecasting*, **18**, 782-794.
- Iacono, M. J., E. J. Mlawer, S. A. Clough, and J.-J. Morcrette, 2000: Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR Community Climate Model, CCM3. *J. Geophys. Res.*, **105**, 14873-14890.
- Janjic, Z. I., 2001: Nonsingular Implementation of the Mellor-Yamada Level 2.5 Scheme in the NCEP Meso Model. *NOAA/NWS/NCEP Office Note #437*, 61 pp.
- Janjic, Z. I., 2003: A nonhydrostatic model based on a new approach. *Meteor. Atmos. Phys.*, **82**, 271-285.
- Janjic, Z. I., T. L. Black, M. E. Pyle, H.-Y. Chuang, E. Rogers, and G. J. DiMego, 2004: The NCEP WRF core. *Preprints, 5th WRF/14th MM5 User's Workshop*, 22-25 June, Boulder, CO, 184-187.
- Kain, J. S., M. E. Baldwin, and S. J. Weiss, P. R. Janish, M. P. Kay, and G. Carbin, 2003a: Subjective verification of numerical models as a component of a

- broader interaction between research and operations. *Wea. Forecasting*, **18**, 847-860.
- Kain, J. S., P. R. Janish, S. J. Weiss, M. E. Baldwin, R. S. Schneider, and H. E. Brooks, 2003b: Collaboration between forecasters and research scientists at the NSSL and SPC: The Spring Program. *Bull. Amer. Meteor. Soc.*, **84**, 1797-1806.
- Kain, J. S., M. E. Baldwin, and S. J. Weiss, 2003c: Parameterized updraft mass flux as a predictor of convective intensity. *Wea. Forecasting*, **18**, 106-116.
- Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *J. Clim. Appl. Met.*, **22**, 1065-1092.
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Met. Mag.*, **30**, 291-303.
- Metz, N. D., D. M. Schultz, and R. H. Johns, 2004: Extratropical Cyclones with Multiple Warm-Front-Like Baroclinic Zones and Their Relationship to Severe Convective Storms. *Wea. Forecasting*, **19**, 907-916.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16663-16682.
- Molinari, J.M., and M. Dudek, 1986: Implicit versus explicit convective heating in numerical weather prediction models. *Mon. Wea. Rev.*, **120**, 326-344.
- Noh, Y., W.G. Cheon, S.-Y. Hong, and S. Raasch, 2003: Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data. *Bound.-Layer Meteor.*, **107**, 401-427.
- Skamarock, W.C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, J. G. Powers, 2005: A Description of the Advanced Research WRF Version 2. *NCAR Tech Note, NCAR/TN-468+STR*, 88 pp. [Available from UCAR Communications, P. O. Box 3000, Boulder, CO 80307].
- Snook, N., and W. A. Gallus, 2004: A Climatology of Severe Weather Reports as a Function of Convective System Morphology Preprints, *22nd Conference on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., CD-ROM, paper P5.5
- Stensrud, D. J., and S. J. Weiss, 2002: Mesoscale model ensemble forecasts of the 3 May 1999 tornado outbreak. *Wea. Forecasting*, **17**, 526-543.
- Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243-1261.
- Tuleya, R. E., 1994: Tropical storm development and decay: Sensitivity to surface boundary conditions. *Mon. Wea. Rev.*, **122**, 291-304.
- Weisman, M. L., J. B. Klemp, and R. Rotunno, 1988: Structure and evolution of numerically simulated squall lines. *J. Atmos. Sci.*, **45**, 1990-2013.
- Weisman, M. L., W. C. Skamarock, and J. B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. *Mon. Wea. Rev.*, **125**, 527-548
- Wilks, D. S., 1995: *Statistical methods in the atmospheric sciences: An introduction*. Academic Press, 467 pp.

APPENDIX. Participants in the 2004 Spring Program

NOAA Agencies:

NWS/NCEP/SPC:

D. Bright
G. Carbin
G. Dial
R. Edwards
S. Goss
J. Levit
C. Mead
J. Peters
J. Racy
S. Weiss

NWS/NCEP/EMC:

J. Du
B. Ferrier
Z. Janjic
G. Manikin
Z. Toth

NWS/NCEP/HPC:

M. Eckert
P. Manousos

NWS/DTX:

G. Mann

OAR/FSL:

J. Brown

Cooperative Institutes:

CIMMS/OU/NSSL:

M. Baldwin
D. Burgess
J. Gourley
J. Kain
D. Schultz
P. Spencer
M. Wandishin

CIRES/CU/CDC:

B. Mapes

Universities:

Arizona:

S. Mullen

Iowa State:

B. Gallus

Missouri-Columbia:

P. Market

Naval Postgraduate School:

R. Elsberry

Oklahoma:

M. Bukovsky

SUNY-Albany:

L. Bosart
T. Galarneau
A. Wasula

Utah:

J. Steenburgh

Valparaiso:

A. French

Wisconsin - Madison:

G. Tripoli

Wisconsin - Milwaukee:

P. Roebber

NRC/NSSL:

V. Homar

Air Force Weather Agency:

R. Curran

S. Rugg

CAPS/OU:

D. Weber

NCAR:

M. Weisman

Meteorological Service of Canada:

B. Bukoski

A. Erfani

P. King

D. Patrick

P. Vaillancourt

Finnish Meteorological Institute:

J. Teittinen

Koch/Entergy Corporation:

P. Janish